# Real_Property_Information: Exploring and Visualizing Real Estate Data in Python

Saad Nadeem, Isaiah Scott, Shuangbao Wang Ph.D.

Computer Science Department, Morgan State University, Baltimore, Maryland

## Introduction

The Real_Property_Information dataset is a collection of real estate data that provides detailed information on properties in Baltimore, including their assessed values, lot sizes, building structures, and ownership information. The dataset is usually maintained by a government agency responsible for assessing property taxes.

## Objectives

- Use the DataFrame API to visualize your data in Google Collab.
- Show measures of central Show box plots for each numerical features or for selected features
- Identify any outliers
- Plot the correlation between each pair of numerical features
- Use the steps at Pythonic Data Cleaning With Pandas and NumPy
- Remove records with missing values or records that you think are inconsistent.
- Chart visualization
- Plot With Pandas
- Develop code that trains, tests and tunes a model to make accurate predictions on the target.

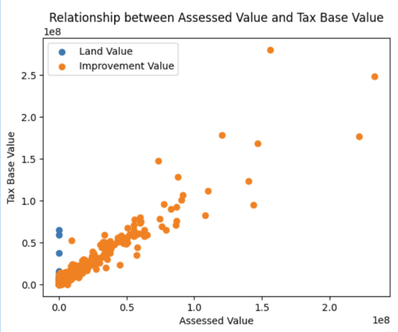
## Visualization with Python

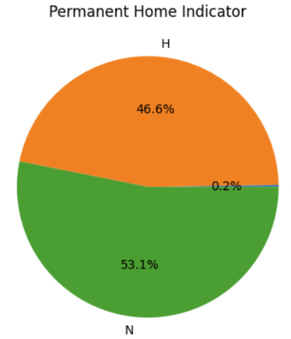
To visualize and relate some of the attributes in the Excel file using Python, we could use various plotting libraries such as Matplotlib, Seaborn, or Plotly. For example, we could create a scatter plot using Matplotlib or Seaborn to visualize the relationship between the "Current full cash" and "Current full cash improvement" attributes. We can also use a scatter plot to visualize the relationship between the "Current full cash" and "Current full cash exempt land" or "Current full cash exempt improvement" attributes. We could use a bar chart to visualize the distribution of properties that are exempt from State and/or local real property taxes based on the "Exempt type city or state" attribute. We could also use a pie chart to show the distribution of different land use code groups based on the "Land use code group" attribute. Alternatively, we could use a bar chart to visualize the frequency of each land use code group. Finally, to relate the "Downtown MGT. district" attribute with other attributes, we could use a bar chart to show the frequency of properties located in the downtown management district based on other attributes such as land use code group or zoning code.

### References and Acknowledgements

https://data.baltimorecity.gov/datasets/real-property-information-2/explore?location=39.284720%2C76.620485%2C12.71&showTable=true

## Results



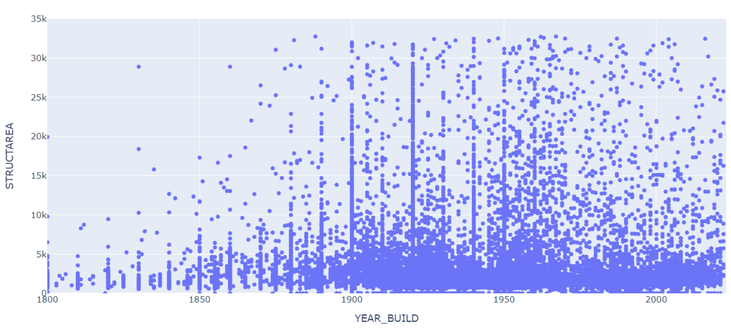

```
MODEL 1 Accuracy:  82.3 %

'\nbest = 0\nfor _ in range(100): # How many generations we want to run this model and save and get bes
(X,Y, test_size=0.1) \n\n  # ====== LINEAR REGRESSION ========\n\n  # Trying a linear regression model
acc = linear.score(x_test, y_test)\n  print(acc)\n\n  # Let's save the best model and work up\n  if acc
pickle.dump(linear, f)\n\npickle_in = open('sectionModel.pickle', 'rb')\nlinear = pickle.load(pickle_in
4),'%')\n# ====================================\n'
```

```
# Let's test it with some predictions
predictions = model.predict(x_test)
print("Accuracy:",metrics.accuracy_score(y_test, predictions))

for x in range(len(predictions)):
  print("\nPrediction: ",predictions[x], "\nInput Data: ",x_test[x], "\nActual Data: ", y_test[x])
# ====================================
```

```
Prediction:  101
Input Data:  (1017, 16, 0, 789, 759, 78, 258, 0, 0, 0, 0, 0)
Actual Data:  101
```




## Conclusions

The Real_Property_Information dataset can be a valuable resource for real estate professionals, policymakers, and researchers interested in the Baltimore City area's housing market. By leveraging the information provided by this dataset and applying advanced analytical techniques, we can gain a better understanding of the real estate market and make informed decisions.